2011-5

The Potential Role of Listening Modes in Auditory Interfaces for Location-based Services

John McGee
Technological University Dublin, johnnyboymcgee@gmail.com

Charlie Cullen
Technological University Dublin, charlie.cullen@tudublin.ie

Recommended Citation
Mcgee, J. & Cullen, C. (2011) The Potential Role of Listening Modes in Auditory Interfaces for Location-based Services. *5th symposium of Forum Klanglandschaft* Florence, Italy, 20–22, May.

THE POTENTIAL ROLE OF LISTENING MODES IN AUDITORY INTERFACES FOR LOCATION-BASED SERVICES

John McGee

Digital Media Centre,
Dublin Institute of Technology,
Aungier Street, Dublin 2, Rep. of Ireland
john.mcgee1@student.dit.ie

Dr. Charlie Cullen

Digital Media Centre,
Dublin Institute of Technology,
Aungier Street, Dublin 2, Rep. of Ireland
charlie.cullen@dmc.dit.ie

ABSTRACT

The auditory modality offers several advantages as a means of communication for the purposes of location-based services (LBS), including fast response time [1], low processing and storage overheads [2], and hands/eyes-free mobility. However, with more and more sound-producing technology being used in day-to-day life, the battle for our acoustic attention has led to a steady rise in acoustic noise levels [3]. In an already noisy environment, it is tempting for the sound designer to simply use more volume as a means of gaining the listener's attention but this only serves to create a vicious circle of noise in which every sound designer is merely struggling to be heard over the noise of every other sound designer. The field of soundscape theory, however, may offer some potential solutions in this regard. Soundscape theory, as described by Schafer [4, 5] and Truax [6], considers sound from a more holistic point of view and the concept of listening modes considers the different levels of attention we pay to auditory stimuli depending on context and location within the soundscape. While several different theoretical listening modes have been proposed across the various acoustic disciplines, there is a need for empirical data to support the existence of these modes. One area in which there is a certain amount of empirical data is in relation to spectral bandwidth and what Krause [7-9] has called his 'niche theory'. Niche theory describes the way in which different species appear to occupy discrete frequency bandwidths within the soundscapes of natural habitats; it is argued that this natural balance keeps redundant noise to a minimum and enables more efficient acoustic communication. If the principles observed in niche theory were to be observed in human listening behaviour, a new approach to sound design might be possible whereby auditory stimuli exploit specific frequency bandwidths in order to maximise information exchange without necessarily raising noise levels. In this paper, we outline a proposed experiment in which listeners are asked to engage in a foreground task that encourages competitive conversation while also attending to a background listening task in which participants have to acknowledge background non-speech sounds of varying spectral bandwidth presented at random intervals. Our aim is to compare the spectrogram information of both the foreground task and the background stimuli to see if relative spectral bandwidth has any discernible effect on stimulus identification success rate and response time.

1. INTRODUCTION

A location-based service (LBS) can be described as an application that is dependent on a certain location [10]. With the increasing capabilities of many modern mobile devices and the increased ubiquity of distributed computing systems in many workplaces and public environments, the emergent field of LBS has come to be of great interest. The auditory modality offers great potential in the field of LBS from the point of view of mobile computing because of the unique advantages it affords over other modalities. For example, audio is hands-free, eyes-free, and focus-independent, it has a faster neural processing rate than visual and haptic stimuli [1], and offers lower storage and processing overheads when compared with visual content [2]. For these reasons the auditory modality has become favourable as a means of communication in situations where the information being delivered is of urgent importance, where the information is of secondary importance and is intended to operate on the periphery of the user's attention (as is the case with ambient interfaces), where bandwidth limitations are an issue, and in situations where the user might require the use of their eyes and/or hands for concurrent tasks. However, one of the major pitfalls facing the sound designer nowadays is the issue of rising acoustic noise levels and the temptation of simply using volume as the primary means of commanding a listener's attention. Some of the theoretical concepts outlined in the field of soundscape theory, particularly the concept of listening modes, may offer some alternative solutions in this regard.

2. THE PROBLEM OF RISING ACOUSTIC NOISE LEVELS

While Western culture has become increasingly demanding in terms of visual attention [11], we at least have the option of choosing where we direct our visual focus; furthermore, if we wish to ignore a visual stimulus altogether we can simply close our eyes. While a similar increase in acoustic noise levels has also occurred in Western society, it is not as easy to ignore auditory stimuli that we do not wish to focus on. The ear canals, in their natural state, are always open and predisposed towards receiving acoustic stimuli. Since the industrial revolution in the late 18th century, the acoustic environment that surrounds us has been getting progressively louder and more cluttered and with each new sound-producing piece of technology that is developed, the battle for our acoustic attention intensifies. Schafer [3] has estimated that this battle for acoustic expression contributes to a rise in environmental sound levels of around 0.5 to 1 decibel per year (figure 1).

3. SOUNDSCAPE THEORY & THE CONCEPT OF LISTENING MODES

Schafer and Truax [6] distinguish between two distinct types

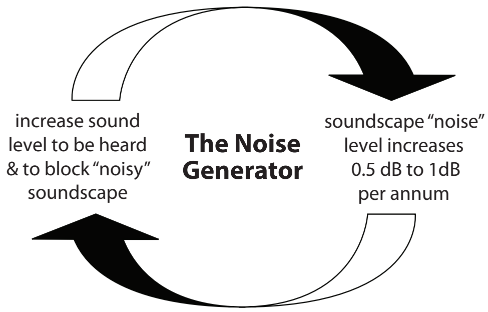

Figure 1: Increasing noise levels (adapted from Wrightson [3]).

of soundscape: hi-fi soundscapes and lo-fi soundscapes. A hi-fi soundscape is defined as any acoustic community in which there is a high degree of information exchange between the elements within it and in which the listener is involved in an interactive relationship with the environment; they are generally associated with rural and non-industrial areas. Hi-fi soundscapes typically have a blend of distinctive and varied acoustic features and discrete sounds are generally heard clearly within them. A lo-fi soundscape, on the other hand, is defined as any unbalanced acoustic community in which there is a high degree of redundancy and a low amount of information exchange; they are commonly associated with modern urban and suburban areas. Lo-fi soundscapes tend to feature numerous competing sounds and poor acoustic definition. Is is often the lo-fi soundscape environment that the sound designer has to consider when approaching the design of an auditory interface for the purposes of LBS.

As well as the different types of soundscapes that surround us, we must also consider the different ways in which we listen to auditory stimuli. While hearing can be regarded as the passive reception of acoustic stimuli by the auditory system, listening, on the other hand, requires a deliberate process through which we interpret the meaning of auditory stimuli [12]. It is quite feasible that two listeners may hear the same auditory stimuli in a shared acoustic environment and yet have completely different listening experiences, the difference depends largely on the mode of listening employed by each listener. Though it may seem obvious that humans do not regard all auditory stimuli equally, there is as of yet no definitive consensus on the number or variety of listening modes employed by humans. There have, however, been a number of broad theoretical models suggested.

3.1. Background Listening

Truax [6] suggests three modes of listening: background listening, listening-in-search, and listening-in-readiness. Background listening refers to the peripheral way in which we process ambient acoustic stimuli (not to be confused with reduced listening, as proposed by Schaeffer [13]). Ambient acoustic stimuli, though often highly redundant in a soundscape, can also provide acoustic background information that gives the listener a high level of environmental awareness and context. Truax [6] regards background listening to be a form of 'distracted listening' and it highlights an interesting feature of the human auditory system in this regard, namely its ability to pull acoustic stimuli in and out of focus depending on context. Though seemingly quite a passive form of listening in its execution, background listening actually requires a sophisticated cognitive process that employs feature detection and pattern recognition, as well as the comparison of said patterns to other known patterns typical to a given environment [14]. Background listening generally requires favourable acoustic conditions and a reasonably good signal-to-noise ratio; it might reasonably be assumed that background listening has become a more prominent mode of listening as a result of steadily rising acoustic noise levels.

3.2. Listening-In-Readiness

Listening-in-readiness is an intermediate form of listening that lies somewhere between background and foreground attention; it describes the process whereby a listener is ready to receive some form of significant auditory stimulus even if their attention is focused elsewhere. Listening-in-readiness relies on associations being built up over time so that sounds can be easily and quickly identified, even via background cognitive processing in the brain. Listening-in-readiness exploits the human auditory system's ability to process both familiar sounds (to determine any deviation from established patterns) and unfamiliar sounds (to allow for the interpretation of any potential significance that these deviations may have) and is closely linked with the mechanics of the 'cocktail party effect' [15, 16]. An example of listening-in-readiness would be a sleeping mother who can ignore the ambient noise of traffic outside but will awaken if she hears her baby crying.

3.3. Listening-In-Search

Listening-in-search, according to Truax [6], is the most consciously involved form of listening. It is a highly active mode of analytical listening whereby the listener scans the acoustic environment searching for auditory cues, ultimately homing in on a specific auditory stream to the exclusion of others. An example of listening-in-search would be the way in which a listener might pay attention to one specific speaker in a conversation or the way in which a blind person listens to the taps of their cane for spatial information.

While several different theoretical listening modes have been proposed across the various acoustic disciplines, there is a need for empirical data to support the existence of these modes. One of the aims of our research is to examine the validity of listening modes under experimental conditions in the hope that specific acoustic factors regarding how auditory information is presented might be identified and subsequently exploited in the design of more efficient auditory interfaces. To this end, we intend to focus initially on the three theoretical listening modes proposed by Truax as they are quite broad in scope and relatively objective in terms of the terminology they employ. Future work may examine more narrowly defined listening modes such as causal, semantic, and reduced listening [17].

4. NICHE THEORY & THE POSSIBLE ROLE OF SPECTRAL BANDWIDTH IN LISTENING MODES

One particular area of soundscape theory where a certain amount of empirical evidence does already exist is in relation to spectral bandwidth and what Krause [7-9] has called 'niche theory'. Niche theory posits that, in natural habitats, coexistent species occupy discrete frequency bandwidths within the local soundscape, resulting in a harmonious biophony in which acoustic information is exchanged efficiently. Krause [7] first observed this phenomenon in the early 1980s while gathering field recordings in Africa; he found that each biome that he captured on tape appeared to exhibit a unique acoustic 'fingerprint' when analysed as a spectrogram and that species appeared to have a clearly defined sense of auditory 'space' as much as they did three-dimensional space (Fig. 2). By comparison, outside of noise by-laws and tacit societal conventions, humans have relatively little regard for any sense of acoustic harmony in a modern urban context. However, if the principles observed in niche theory were to be observed in human listening behaviour then a new approach to sound design might be possible whereby auditory stimuli exploit specific frequency bandwidths in order to maximise information exchange without necessarily raising noise levels.

5. EXPERIMENTAL DESIGN

We propose an experiment aimed at exploring what effect spectral bandwidth might have in relation to the perception of non-speech background stimuli in the presence of competitive foreground conversation. If the findings of Krause [7-9] are anything to go by, one might assume two possible outcomes when utilising background stimuli that occupy specific targeted bandwidths in a competitive acoustic scenario: (1) that background stimuli occupying a bandwidth which does not significantly overlap with that of foreground stimuli will be perceived as more salient and thus fare better in registering the attention of test subjects, or (2) that background stimuli occupying a bandwidth which does not significantly overlap with that of foreground stimuli will be perceived as an integrated part of the keynote sound (i.e. the aggregate of all low-level sounds heard as a background texture and not as distinct, individual acoustic components [6]) and thus fare less successfully in registering the attention of test subjects. The null hypothesis in this case would be that spectral bandwidth has no significant effect either way on how non-speech background stimuli are perceived within a competitive acoustic scenario. A third assumption within this context might be that, given the human auditory system's apparent predilection for the frequency range involved in human speech [18], background stimuli operating between 1 kHz and 4 kHz (the approximate range of the important resonance frequencies of the human vocal tract, which determine the acoustic character of a speech sound) may fare better in registering the attention of test subjects than stimuli operating outside of this range.

5.1. Methodology

Pairs of test subjects are placed in separate sound isolation booths (each fitted with a microphone, a foot-switch, and a computer monitor) and asked to converse with one another over headphones in order to solve a shared puzzle presented on their computer screens within a five minute time limit. Along with this foreground task, each test subject is also presented with background stimuli consisting of simple non-speech environmental sounds of varying spectral bandwidths at random intervals. Subjects are informed prior to beginning the experiment that their primary task is to solve the shared puzzle but that they may also hear background noises, which they are to acknowledge by pressing on the foot-switch. The dialogue of both test subjects will be recorded along with their background stimuli tracks and timestamped foot-switch responses. On completion of the experiment, subjects will be asked to fill out

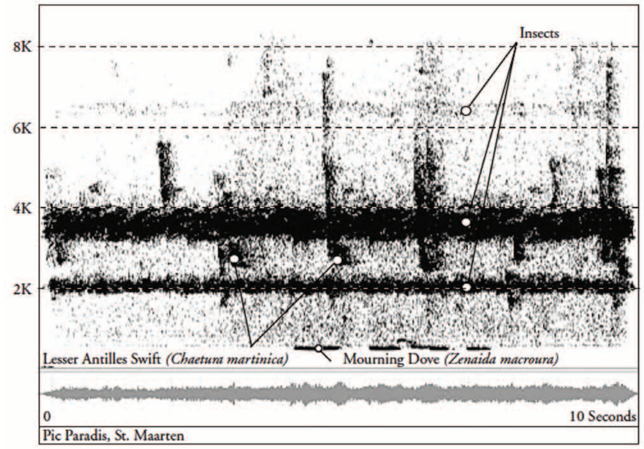

Figure 2: Field recording spectrogram showing different species and their relative space within the spectral bandwidth (taken from Krause [7]).

a NASA TLX survey [19] to assess their perception of how the background task affected their performance of the foreground task.

5.2. Experimental Conditions

In order to limit the effects of extraneous auditory stimuli, In order to limit the effects of extraneous auditory stimuli, experiments will take place in soundproof isolation booths using Beyerdynamic DT 150 headphones with a nominal frequency response of 5 - 30,000 Hz, a nominal SPL of 97 dB, and ambient noise attenuation of approximately 20 dBA. A Pro Tools HD7 rig and high-quality microphones will be used to record all of the audio from the experiment at 192 Khz/24-bit (Fig. 3). Test subjects will each use a Boss FS-5L latch foot switch to acknowledge background stimuli during the experiment. When activated, the Boss FS-5L latch foot switch registers a clearly defined spike in the recorded waveform, which offers sub-second accuracy and will be used to record test subjects' reaction time. Each booth will contain a computer monitor presenting a common view of the shared foreground task, the GUI of which will be manipulated by the experimenter according to the verbal instructions of the two test subjects.

5.3. Foreground Task Design

In order to generate competitive conversation, participants are presented with one of three imaginary survival scenarios to discuss as a foreground task: a shipwreck, being stranded on the moon, and being stranded on a mountain. This approach has been used successfully in the past to elicit natural and spontaneous conversation between test subjects [20] and to observe speech convergence in natural dialogue [21]. The monitor in each isolation booth will present the participants with a shared view of 8 items, which they are to arrange in order of importance for their survival. The experimenter, monitoring the conversation through his own headphones, will arrange the order of the objects in accordance with the consensus opinion of the test subjects. The time limit of 5 minutes is intended to impose a sense of urgency on proceedings and thus encourage competitive conversation.

Figure 3: Overview of experimental design setup

5.4 Background Task Design

Along with the foreground task, participants are instructed that they may hear noises in the background, which they are to acknowledge by pressing on their foot switch. These background sounds will consist of short non-speech environmental sounds of varying spectral bandwidth presented at random intervals throughout the experiment. Test subjects will each be presented with their own unique sequence of background stimuli so as to discourage them from discussing the background task with each other when the experiment is in progress. These background sounds will be manipulated beforehand via bandwidth filtering to ensure they occupy specific targeted bandwidths; aside from the frequency bandwidth, all other acoustic variables such as volume, timbre, duration, and so on, will be kept constant.

5.5. Sample Selection

The focus of this research is the listening behaviour of the normal adult population and so for this reason test subjects will be between 18-65 years of age and be of healthy hearing. Age and gender will be documented with the names of participants redacted to protect privacy.

5.6. Prospective Outcomes

The NASA TLX survey will be used to assess what cognitive load is placed on test subjects as a result of the concurrent foreground and background tasks. A control condition in which there is no background task should reveal any difference in perceived cognitive load between the test condition and the control condition. The recordings of the two foreground conversation tracks along with the two background stimuli tracks and timestamped foot switch tracks will give us three specific data sets. Firstly, they will identify whether or not the subjects were successful in identifying the background stimulus tones when they were presented. Secondly, they will allow us to calculate the subjects' relative response time in acknowledging the background stimuli in the event that they were successful in detecting them. Finally, if we analyse the spectrogram information of both the foreground conversation tracks and background stimuli tracks we will be able to assess what spectral bandwidth they occupy in relation to each other and compare this with success rate and response time to see if their is any positive or negative correlation.

6. CONCLUSIONS

In this paper, we discussed the emergent field of location-based services and the advantages that audio offers as a modality in this regard. One of the main challenges facing sound designers is the issue of rising acoustic noise levels and the tendency to use volume as the primary means of gaining a listener's attention. However, there are a number of theories from the field of soundscape theory which may offer potential solutions to these problems, particularly the concept of listening modes. There is still a need for more empirical data with which to apply these listening modes in practice. With this in mind, taking Krause's niche theory as a starting point, we have outlined the design of an experiment aimed at investigating what effect spectral bandwidth might have in relation to the perception of non-speech background stimuli in the presence of competitive foreground conversation.

9. REFERENCES

[1] R. Kail and T. A. Salthouse, "Processing Speed as a Mental Capacity," Acta Psychologica, vol. 86, pp. 199-225, August 1994.

[2] R. McGee, "Auditory Displays and Sonification: Introduction and Overview," In press. 2009. Available: http://www.

lifeorange.com/writing/Sonification_Auditory_Display.pdf

[3] K. Wrightson, "An Introduction to Acoustic Ecology," Soundscape: The Journal of Acoustic Ecology, vol. 1, pp. 10-13, Spring 2000.

[4] R. M. Schafer, The New Soundscape: a Handbook for the Modern Music Teacher. Ontario: Don Mills, 1969.

[5] R. M. Schafer, The Tuning of the World, 1st ed. New York: Random House Inc., 1977.

[6] B. Truax, Acoustic Communication, 2nd ed. Westport, Connecticut: Ablex Publishing, 2001.

[7] B. L. Krause, "The Niche Hypothesis: How Animals Taught Us to Dance and Sing," Bioacoustics, Habitat Ambience in Ecological Balance, p. 6, Winter 1987.

[8] B. L. Krause. (1993, June) The Niche Hypothesis: A Virtual Symphony of Animal Sounds, The Origins of Musical Expression and the Health of Habitats. The Soundscape Newsletter. 6-10. Available: http://interact.uoregon.edu/Medialit/wfae/library/newsletter/SNL6.PDF

[9] B. L. Krause. (1997, Winter) The Niche Hypothesis: Creature Vocalizations and the Relationship Between Natural Sound and Music. Nature Sounds Newsletter. Available: http://www.naturesounds.org/Newsletters/Win97NicheHypothesis.html

[10] T. D'Roza and G. Bilchev, "An Overview of Location-based Services," BT Technology Journal, vol. 21, pp. 20-27, January 2003.

[11] M. Rampley, Exploring Visual Culture: Definitions, Concepts, Contexts. Edinburgh: Edinburgh University Press, 2005.

[12] G. Ferrington, "Soundscape Studies: Listening with Attentive Ears," Telemedium, The Journal of Media Literacy, vol. 49/50, pp. 42-45, Spring 2003.

[13] P. Schaeffer, Traité des Objets Musicaux (Treatise on Musical Objects). Paris: Le Seuil, 1966.

[14] B. Truax. (1995, 1/11/09). Sound in Context: Acoustic Communication and Soundscape Research at Simon Fraser University. 1-11. Available: http://interact.uoregon.edu/MediaLIt/wfae/library/articles/truax_SFUniversity.pdf

[15] B. Arons, "A Review of the Cocktail Party Effect," Journal of the American Voice I/O Society, vol. 12, pp. 35-50, 1992.

[16] S. Haykin and Z. Chen, "The Cocktail Party Problem," Neural Computation, vol. 17, pp. 1875-1902, September 2005.

[17] M. Chion, "The Three Listening Modes," in Audio-vision - Sound on Screen, C. Gorbman, Ed., ed New York: Columbia University Press, 1994, pp. 25-34.

[18] B. DeBoer, "The Evolution of Speech," in Encyclopedia of Language and Linguistics, K. Brown, Ed., 2nd edition ed: Elsevier, 2006, p. 11000.

[19] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in Human Mental Workload, P. A. Hancock and N. Meshkati, Eds., ed Amsterdam: Elsevier, 1988, pp. 139-183.

[20] M. Rehm, et al., "The CUBE-G Approach – Coaching Culture-specific Nonverbal Behavior by Virtual Agents," in In Proceedings of the 38th Conference of the International Simulation and Gaming Association (ISAGA'07), Nijmegen, The Netherlands, 2007.

[21] S. Kousidis, et al., "Time Series Analysis of Acoustic Feature Convergence in Human Dialogues," in SPECOM 2009, St. Petersburg, Russian Federation, 2009.